Atlas LSH Neural Networks: Routing Granularity as a Design Axis

Executive Summary

Atlas LSH neural networks perform input-dependent parameter selection using locality sensitive hash (LSH) codes.

A central architectural question is:

> At what granularity should routing decisions select parameters?

The design space spans a spectrum from whole-matrix selection to elementwise parameter selection. Different points along this spectrum are expected to exhibit different trade-offs between capacity, parameter sharing, specialization, and generalization.

Routing Granularity Spectrum

Consider an n x n weight matrix.

1. Whole Matrix Selection

Store multiple complete matrices:

$W^{(0)}$, $W^{(1)}$, ...

LSH bits select one matrix.

$W(x)=W^{(b(x))}$

Characteristics

- Very coarse routing.

- Few routing bits required.
- Extremely high parameter reuse.
- Large overlap between computational routes.
- Strong gradient sharing across examples.
- Smooth generalization behavior.
- Limited local adaptation.

Expected Behavior

- Robust learning.
- Good statistical efficiency.
- Lower risk of over-specialization.
- May underfit complex problems.

---

2. Block-Level Selection

Partition matrices into blocks:

W=

$|B_{11}\ B_{12}|$
$|B_{21}\ B_{22}|$

LSH bits select:

- row blocks,
- column blocks,
- tiles,
- channels,
- attention heads.

Characteristics

- Intermediate routing granularity.
- Moderate parameter reuse.
- Moderate route overlap.
- Increased specialization.
- Retains substantial gradient sharing.

Expected Behavior

- Good balance between flexibility and generalization.
- Increased capacity relative to whole-matrix routing.
- Likely to provide a useful operating regime for many problems.

---

3. Elementwise Binary Selection

Each matrix element contains two candidate values:

$$w_{ij} \in \{w_{ij}^{(0)}, w_{ij}^{(1)}\}$$

One LSH bit selects each element.

$$w_{ij}^{(x)} = w_{ij}^{(bij(x))}$$

Requires:

$$n^2$$

routing bits.

The number of possible matrices becomes

$$2^{nn}$$

Characteristics

- Fine-grained routing.
- Extremely large operator family.
- High local adaptability.
- Reduced parameter sharing.
- Reduced route overlap.

Expected Behavior

- Fine quantization of parameter space.
- Strong local specialization.
- Potentially improved approximation capability.
- Generalization may remain good provided neighboring inputs share many routing bits.

---

4. Elementwise Multiway Selection

Each matrix element selects from

2^k

candidate values:

$w_{ij} \in \{w_{ij}(0), \dots, w_{ij}(2^{k-1})\}$

requiring

kn^2

routing bits.

The number of possible matrices becomes

2^{knn}.

Characteristics

- Extremely fine routing granularity.
- Vast combinatorial operator space.
- Very high routing entropy.
- Low parameter overlap between routes.
- Strong specialization.

Expected Behavior

- Highly quantized function space.
- Reduced gradient sharing.
- Increased risk of route fragmentation.
- Increased risk of memorization.
- Generalization may degrade if neighboring inputs no longer share sufficient parameters.

A Possible Behavioral Spectrum

Moving from coarse to fine routing:

Whole matrix
↓
Block selection
↓
Elementwise binary selection
↓
Elementwise multiway selection

Expected trends:

- Capacity increases.
- Specialization increases.
- Routing entropy increases.
- Parameter reuse decreases.
- Gradient sharing decreases.
- Route overlap decreases.

Possible progression:

Too coarse
↓
Underfitting

Intermediate granularity
↓
Best generalization

Extremely fine granularity
↓
Over-specialization
↓
Memorization

Conclusion

Atlas LSH neural networks introduce routing granularity as a new architectural design axis.

The central trade-off is between:

- representational capacity,
- specialization,
- parameter reuse,
- route overlap,
- and generalization.

The optimal operating regime may lie between the extremes of coarse whole-matrix routing and fully independent elementwise routing.